PART B - (5 * 13 = 65 marks)

11. a. i. Compare and contrast Supervised learning, Unsupervised learning, semi-supervised learning and reinforcement learning. (9)

ii. A manufacturer claims that its drug test will detect steroid use (shows positive for an athelete who uses steroids) 95% of the time. What the company doesn't tell you in that 15% of all the steroid-free individuals also test positive (false positive rate). 10% of the rugby team members uses steroids. Your friend on the rugby team has just tested positive. Find the probability that he uses steroids. Assume E – then event that a rugby team member tests positive and F – the event that the rugby team member uses steroids. (4)

(OR)

11. b. i. What are the three components of Decision theory? Explain the three ways in which decision theory problems are represented? (3)

ii. You are a robot in a lumber yard and must learn to discriminate Oak wood from Pine wood. You choose to learn a Decision tree classifier. You are given the following table:

Example	Density	Grain	Hardness	Class
1	Heavy	Small	Hard	Oak
2	Heavy	Large	Hard	Oak
3	Heavy	Small	Hard	Oak
4	Light	Large	Soft	Oak
5	Light	Large	Hard	Pine
6	Heavy	Small	Soft	Pine
7	Heavy	Large	Soft	Pine
8	Heavy	Small	Soft	Pine

1. Which attribute would information gain choose as the root of the tree? (2)
2. Draw the decision tree that would be constructed by recursively applying the information gain to select roots of the sub-trees. Show the steps. (4)
3. Classify these new examples as Oak or Pine using the constructed decision tree. (4)

   a. [Density = Light, Grain = Small, Hardness = Hard]

   b. [Density = Light, Grain = Small, Hardness = Soft]

12. a. i. For the three objects, A: (1, 0, 1, 1), B: (2, 1, 0, 2) and C: (2, 2, 2, 1) store them in a data matrix and use Manhattan, Euclidean and Cosine distances to generate the respective distance matrices. (6)

ii. A person doesn't feel well and goes to a doctor. Assume two states of nature:

- w1: The person has a common flu
- w2: The person is really sick (a vicious bacterial infection)

The doctors prior probabilities are: p(w1) = 0.9 and p(w2) = 0.1. Although this doctor can diagnose with very high rate of success using the prior, he/she can lose a patient once in a while. Denote the two possible actions: a1 = prescribe hot tea and a2 = prescribe antibiotics. Assume the following cost (loss) matrix.

		w1	w2
$\lambda_{i,j}$	a1	0	10
	a2	1	0

Assume the doctor also performs a blood test. Suppose the possible results of the blood test are x1 = negative (no bacterial infection) and x2 = positive (infection) and the constraint is that blood tests can often fail. If the class conditional probabilities are

- infection, p(x1 | w2) = 0.3, p(x2 | w2) = 0.7
- flu, p(x2 | w1) = 0.2, p(x1 | w1) = 0.8

Compute the conditional risk for each action and observation so that the doctor can choose an optimal action that minimizes risk. (7)

(OR)

12. b. i. How is regression analysis used in Machine learning. (4)

ii. Explain the taxonomy of three methods of classification, listing at least two methodologies under each. Discuss the Pros and Cons of each method listed by you. (9)

13. a. i. Given the following data:

Object	Attribute 1 (x)	Attribute 2 (y)
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Cluster the four objects with (x,y) representing locations into two clusters. Initial cluster centres are: Medicine A (1, 1) and Medicine B (2, 1). Use Euclidean distance and k-means algorithm to find the two cluster centres after the first iteration. (7)

ii. Explain the PCA method of Dimensionality reduction. (6)

(OR)

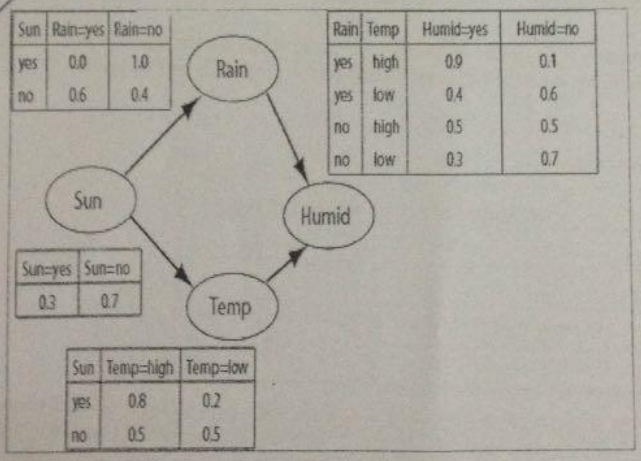
13. b. i. We have two coins: **A** and **B** whose probabilities for heads are q_A & q_B respectively. 6 measurement sets are taken with 15 coin tosses in each set. Calculate Maximum Likelihood probabilities for q_A & q_B knowing which coin is tossed for each set. Use EM algorithm and show **TWO** iterations for the following data: (8)

Coin Type	6 sets of 15 tosses each	Coin A	Coin B
B	H T T T H H T H T H T H H T H		
A	[illegible]		
A	[illegible]		
B	H T H T T T H H T T T T T T T		
A	T H H H T H H H T H H H H H H		
B	T H T H T H T H T T T H H H T		

ii. Discuss Bagging and Boosting in brief. (5)

14. a. i. What are probabilistic graphical models? Discuss. (4)

ii. Consider the following Bayes belief network containing four Boolean random variables.



Sun	Rain=yes	Rain=no
yes	0.0	1.0
no	0.6	0.4

Rain	Temp	Humid=yes	Humid=no
yes	high	0.9	0.1
yes	low	0.4	0.6
no	high	0.5	0.5
no	low	0.3	0.7

Sun=yes	Sun=no
0.3	0.7

Sun	Temp=high	Temp=low
yes	0.8	0.2
no	0.5	0.5

i. Find P(S = T, Rain = Yes, Temp = low, Humid = Yes) (3)

ii. Given (Sun = T), find P(Humid = Yes). (3)

iii. Given (Humid = F), find P(Sun). (3)

(OR)

14. b. i. Discuss briefly Hidden Markov model indicating the various probabilities that are associated with it. (4)

ii. Ram is a three-month old baby. He can be happy, hungry or having a wet-diaper. Initially when he wakes up from his nap at 1 pm, he is happy. If he is happy, there is a 50% change that he will remain happy one hour later, a 25% chance to be hungry by then and a 25% chance to have a wet diaper. Similarly if he is hungry, one hour later he will be happy with a 25% chance, hungry with 25% chance and wet diaper with 50% chance. If he has a wet-diaper, one hour later he will be happy with 50% chance, hungry with 25% chance, and wet-diaper with 25% chance. When he is happy, he smiles 75% of the time and cries 25% of the time; when he is hungry, he smiles 25% and cries 75% and when he has a wet-diaper he smiles 50% and cries 50%.

- Draw the HMM that corresponds to the above story indicating the transition and output probabilities. (6)
- The care-taker of the child left a note to Ram's parents: "1 pm: smile, 2 pm: cry, 3pm: smile". What is the probability that this particular observed sequence happens? (3)

15. a. i. Discuss the necessity for Sampling and explain any one sampling technique. (5)

ii. Discuss the three general areas of COLT and explain the PAC learning framework. (8)

(OR)

15. b. i. Discuss the elements of Reinforcement learning with an example. (7)

ii. Differentiate between Passive and Active Reinforcement learning. (6)

PART C - (1 * 15 = 15 marks)

16. Choose an **application that is not trivial** and use the same for the following sub-sections and answer them:

a. Explain the use of machine learning for an application of your choice indicating (7)

- Details of the application including assumptions
- How a human being would tackle the application
- Type of machine learning problem
- The input and expected output
- Features to be used
- Possible evaluation strategy to be used to measure

b. Discuss in detail the key steps of machine learning using as illustration the example discussed above. (5)

c. Suggest possible learning technique that can be used for the same example and justify your choice. (3)

Department of Computer Technology, MIT Campus, Anna University

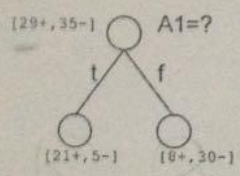
CP7253 - MACHINE LEARNING TECHNIQUES

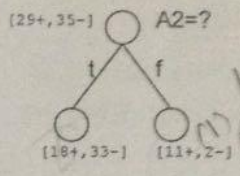
2/4 MECSE – Assessment-I

Date: 09/03/17 Duration:1 hrs Marks:50

Part-A **(4 X 2=8)**

1. Write short notes on design of a learning systems.
2. Which attribute is the best classifier, Justify





3. How to minimize the misclassification rate in decision theory?
4. Weather quadratic regularization penalty is added for the parameter. When estimating a logistic regression model ensures that some of the parameter(weight associated with the components of the input vectors) vanish. Justify your answer.

Part-B **(3 X 9= 27)**

(Answer any three only)

5. There are two coins C1 and C2. C1 has a equal prior on a head (H=1) or tail (T=0) and the fate of C2 is dependent on C1. If C1 is a head, C2 will be a head with probability 0.7. If C1 is a tail, C2 will be a head with probability 0.5. C1 and C2 are tossed in sequence once, and the observed sum of the two coins, S = C1 + C2, is 1. What is the probability that C1=T and C2=H (*Hints: use Bayes theorem*)?
6. Discuss the least square approach for linear model for regression and its relation to maximum likelihood for minimizing the noise ratio relate to their target data set in detail.
7. We are interested here in a particular 1-dimensional linear regression problem. The dataset corresponding to this problem has n examples$(x_1,y_1),\ldots,(x_n,y_n)$, where xi and y_i are real numbers for all i. Part of the difficulty here is that we don't have access to the inputs or outputs directly. We don't even know the number of examples in the dataset. We are, however, able to get a few number computed from the data.

   Let $W^* = [\omega_0^* \ \omega_1^*]^T$ be the least squares solution we are after. In other words, w^* minimizes

$$J(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \omega_0 - \omega_1 x_i)^2$$

   You can assume for our purpose here that the solution is unique.

There are several numbers (statistic) computed from the data that we can use to infer w*. These are

$$\bar{x} = \frac{1}{n}\sum_{k=0}^{n} y_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad C_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$C_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \quad C_{yy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

i. Suppose we only care about the value of ω_1^*. We'd like to determine ω_1^* on the basis of only two numbers (statistics) listed above. Which two number do we need for this?

8. How to learn the parameter function resides over the discriminant function for linear classification. Discuss in detail with suitable illustration.

Part-C **(1 X 15= 15)**

(Answer all the questions)

Rec ID	Age	Income	Student	Credit_ Rating	Buy_ Computer
1	Young	High	No	Fair	--
2	Young	High	No	Excellent	--
3	Medium	High	No	Fair	+
4	Old	Medium	No	Fair	+
5	Old	Low	Yes	Fair	+
6	Young	Low	Yes	Excellent	--
7	Medium	Low	Yes	Excellent	+
8	Young	Medium	No	Fair	--
9	Young	Low	Yes	Fair	+
10	Old	Medium	Yes	Fair	+
11	Young	Medium	Yes	Excellent	+
12	Medium	Medium	No	Excellent	+
13	Medium	High	Yes	Fair	+
14	Old	Medium	No	Excellent	--

9. Will a young student with medium income and fair credit rating buy a computer?

v) Compute the prior probability for each class

vi) Compute the probability of each attribute value given each class

vii) Compute the likelihood of instance z given each class

viii) Find the most probable class

Department of Computer Technology, MIT Campus, Anna University

CP7253 - MACHINE LEARNING TECHNIQUES

2/4 MECSE – Assessment-II

Date: 30/04/17 Duration:1.30 hrs Marks:50

Part A (5 X 10 = 50)

1. Explain briefly about iterative optimization technique. Discuss in detail about an Expectation Maximization algorithm for Gaussian mixture model and maximum likelihood.
2. Explain with suitable illustration about Hidden markov model with transition and observation probability.
3. Support vector machine work by finding a "maximum margin" separating hyperplane in a high-dimensional feature space.
   a) Why is the maximum margin hyperplane better then other separting hyperplanes?
   b) Support vector machines use a special trick for computing scalar product efficiently when fitting a hyperplane in a high-dimensional feature space. What is the trick?
   c) Consider a 2 class classification problem with a dataset of inputs $\{x_1 = (-1, -1), x_2 = (-1, +1), x_3=(+1,-1), x_4 = (+1, +1)\}$ and a corresponding set of targets {t1, t2, t3, t4} where $t_i \varepsilon \{+1, -1\}$. Sing this feature space (no kernel trick), can we build a SVM to perfectly classify this dataset regardless of values of t_i's?
4. Consider the data about student given in the table below. The only attributes are age and three marks. Find the suitable method that has converged the cluster membership of the student for the given dataset.

Student	Age	Mark1	Mark2	Mark3
S_1	18	73	75	57
S_2	18	79	85	75
S_3	23	70	70	52
S_4	20	55	55	55
S_5	22	85	86	87
S_6	19	91	90	89
S_7	20	70	65	60
S_8	21	53	56	59
S_9	19	82	82	60
S_{10}	47	75	76	77
S_{11}	28	73	75	59
S_{12}	18	69	85	55
S_{13}	23	65	50	67
S_{14}	20	85	55	95
S_{15}	32	85	96	84

Training data for a classification problem:

Name	Eggs	Pouch	Files	Class
Cockatoo	Yes	No	Yes	Bird
Dugong	No	No	No	Mammal
Eagle	Yes	No	Yes	Bird
Echidna	Yes	Yes	No	Marsupial
Emu	Yes	No	No	Bird
Flying Fox	No	No	Yes	Mammal
Kangaroo	No	Yes	No	Marsupial
Koala	No	Yes	No	Marsupial
Kookaburra	Yes	No	Yes	Bird
Owl	Yes	No	Yes	Bird
Penguin	Yes	No	No	Bird
Platypus	Yes	No	No	Mammal
Possum	No	Yes	No	Marsupial
Wallaby	No	Yes	No	Marsupial
Whale	No	No	No	Mammal
Dolphin	No	No	No	Mammal

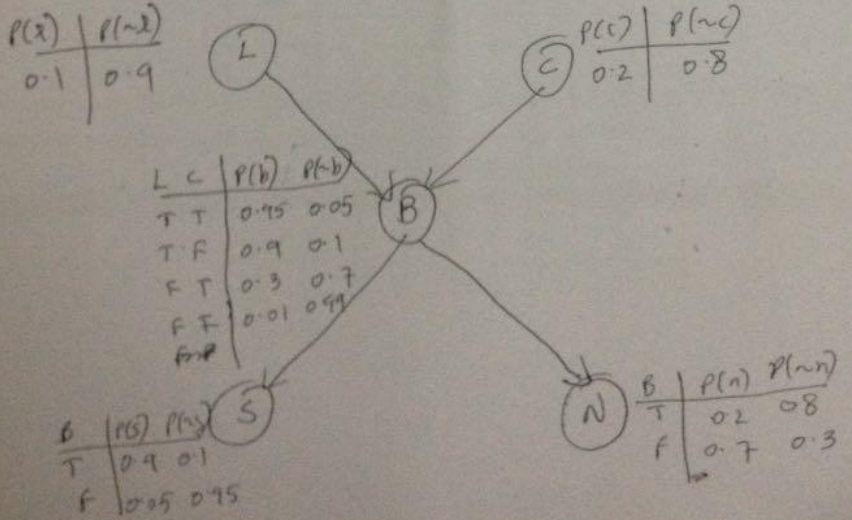
B=6
M=5
MS=3
16

P(E|B)
P(E|M)

16

5. Will all the eggs are bird.
   i) Compute the prior probability for each class
   ii) Compute the probability of each attribute value given each class
   iii) Compute the likelihood of instance z given each class
   iv) Find the most probable class

6. Compute the probability of taking notes (N) in the Bayesian network as shown below. The problem describes the chance of you taking notes in the lecturer or sleeping (S) according to whether or not the course was bouring (B) based on whether or not the professor is boring (L) and the content is dull (C). compute the change of falling asleep in a lecture given that the professor and the course are boring.



Anna University, MIT Campus
Department of Computer Technology
CP 7253 Machine Learning Techniques

Date: 4.03.2016 Time: 9.00 AM– 10.30 AM
Class: M.E.,(CSE) 2/4 Marks: 50

Part A 4 x 3 =12

1. Consider the following 4 training examples
   - x = −1, y = 0.0319
   - x = 0, y = 0.8692
   - x = 1, y = 1.9566
   - x = 2, y = 3.0343

   We want to learn a function f(x) = ax+b which is parametrized by (a, b). Using squared error as the loss function, which of the following parameters would you use to model this function.
   (a) (1, 1)
   (b) (1, 2)
   (c) (2, 1)
2. A medical company touts its new test for a certain genetic disorder. The false negative rate is small: if you have the disorder, the probability that the test returns a positive result is 0.999. The false positive rate is also small: if you do not have the disorder, the probability that the test returns a positive result is only 0.005. Assume that 2% of the population has the disorder. If a person chosen uniformly from the population is tested and the result comes back positive, what is the probability that the person has the disorder?
3. What is pruning? What is the advantage of pruning?
4. Define Expectation and covariance of a function.

Part B 1 X 14 =14

5. a) Explain the Bias- Variance trade off in the context of linear basis function models. (10)
   b) What is regularization and discuss the need for regularization? (4)

Part C 3 X 12 = 24

6. i) a) Construct a decision tree to decide on the allotement of quarters for staff based on rank and experience using the given training sample (8)

Name	Rank	Years of Experience	Allotment of Quarters
A	Asst. Prof	3	No
B	Asst. Prof	7	Yes
C	Prof.	2	Yes

D	Asso Prof	7	Yes
E	Asst. Prof	6	No
F	Asso. Prof	3	No

b) What is machine learning? Explain its types with example. (5)

(OR)

ii) Explain the parameter selection process for linear regression model based on the maximum likelihood and least squares.

7. i) a) Derive the Fisher's discriminant function for multiple classes. (8)

b) Write short notes on boosting. (5)

(OR)

ii) A researcher wants to know if there is a relationship between the number of shopping centers in a state and the retail sales (in billions Rs.) of that state. A random sample of 8 states is listed below. Using the given data answer the following questions.

State	No. of shopping Centers	Sales
1	630	15.5
2	370	7.5
3	616	13.9
4	700	18.7
5	430	8.2
6	568	13.2
7	1200	23.0
8	2976	87.3

a. Interpret the slope in the words of the problem.

b. What is the equation of the regression line?

c. Find the error for predicting the sales of a state with 1200 stores.

"Write what you know. That should leave you with a lot of free time."

— Howard Nemerov

M.E. (Computer Science and Engineering)

III Semester

Machine Learning Techniques

Assessment Test II

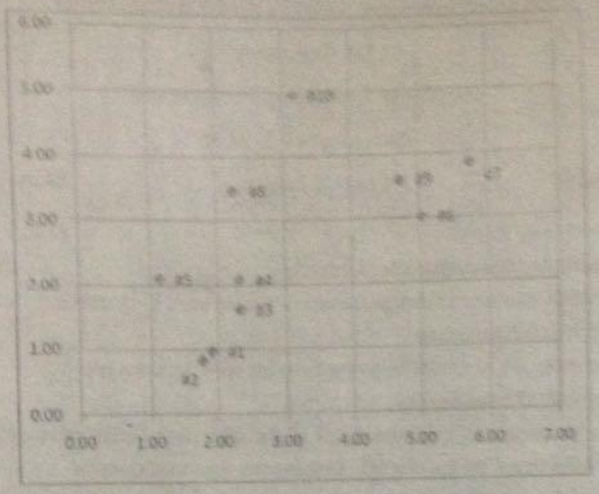
Date : 25-03-2014 Time:2 hours Marks:50 marks

1. Explain briefly any four (4X2 = 8 marks)
   a. Criteria for Dimensionality Reduction
   b. Describe the evaluation parameter - Purity used in clustering.
   c. Bagging and Boosting
   d. Hamming distance for binary attributes with an example
   e. Curse of Dimensionality
2. Explain the taxonomy of three methods of classifications listing at least two methodologies listed under of each. Discuss the Pros and Cons of each method. (8 marks)
3. Briefly explain Least Squares method for learning the parameters of a linear discriminant function use for Classification. (4 marks)
4. Discuss the steps to show how entropy is used to build the decision tree for the following example (10 marks)

Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*Wait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

5. Suppose you are given the following <x,y> pairs. You will simulate the k-means algorithm to identify TWO clusters. Suppose you are given initial assignment cluster center as {cluster1: #1}, {cluster2: #10} – the first data point is used as the first cluster center and the 10-th as the second cluster center. Please simulate the k-means (k=2) algorithm for ONE iteration. What are the cluster assignments after ONE iteration? Assume k-means uses Euclidean distance. What are the cluster assignments until convergence? (Fill in the table below) (8 marks)

Data #	Cluster Assignment after One Iteration	Cluster Assignment after convergence



Data #	X	Y
1	1.90	0.97
2	1.76	0.84
3	2.32	1.63
4	2.31	2.09
5	1.14	2.11
6	5.02	3.02
7	5.74	3.84
8	2.25	3.47
9	4.71	3.60
10	3.17	4.96

6. Explain with the help of a mathematical equation the activation and basis functions of Feed Forward neural networks. (6 marks)
7. Explain the PCA method of Dimensionality reduction. (6 marks)

Department of Computer Technology, MIT Campus, Anna University
CP8202-MACHINE LEARNING TECHNIQUES
Assessment-II

Class: M.E -CSE(2/4) **Date:17.03.2014**
Total: 25 Marks **Time:1 Hour**

PART-A (3x2=6)

1. Which distribution model is used for continuous variables? Give its form.
2. Why we go for Fishers Linear discriminant model? Give its limitations.
3. Does adding Regularization term to an error function controls over fitting? Justify your answer.

PART-B (7x3=21 marks)

1. Describe Linear basis function models with neat diagrams. (5 marks)
2. Present the discriminant functions for Linear classification models. How to learn the parameter for these functions? (7 marks)
3. Write the K-means clustering algorithm. What are the limitations of it? Consider the following eight points P1(2,2), P2(1,14), P3(10,7), P4(1,11), P5(3,4), P6(11,8), P7(4,3), P8(12,9). Take P1,P2,P7 as initial centroids. Calculate two successive positions of those centroids. (7 marks)

**** "Life is like a flowing river, full of opportunities it's up to you whether you take bucket or a spoon from it" ****

Department of Computer Technology, MIT Campus, Anna University
CP8202-MACHINE LEARNING TECHNIQUES
Assessment-II

Class: M.E -CSE(2/4) **Date:17.03.2014**
Total: 25 Marks **Time:1 Hour**

PART-A (3x2=6)

1. Which distribution model is used for continuous variables? Give its form.
2. Why we go for Fishers Linear discriminant model? Give its limitations.
3. Does adding Regularization term to an error function controls over fitting? Justify your answer.

PART-B (7x3=21 marks)

4. Describe Linear basis function models with neat diagrams. (5 marks)
5. Present the discriminant functions for Linear classification models. How to learn the parameter for these functions? (7 marks)
6. Write the K-means clustering algorithm. What are the limitations of it? Consider the following eight points P1(2,2), P2(1,14), P3(10,7), P4(1,11), P5(3,4), P6(11,8), P7(4,3), P8(12,9). Take P1,P2,P7 as initial centroids. Calculate two successive positions of those centroids. (7 marks)

**** "Life is like a flowing river, full of opportunities it's up to you whether you take bucket or a spoon from it" ****

Anna University: Chennai - Madras Institute of Technology
Department Of Computer Technology
CS 8601 Artificial Intelligence
Assessment-3

Class : 6/8 CSE R&S batch
Duration : 1 Hr
Date : 06/04/2015
Max. Marks : 25

PART-A (5x2=10)

1. What are Categories and Objects? (2)
2. What is situation calculus? How to solve the representational frame problem? (2)
3. How does an SVM find the maximum marginal hyper plane and the support vectors?(2)
4. Compare information extraction with information retrieval. (2)
5. In the given Prolog code, what is the value of X? Justify. (2)

```
likes(john,flowers).
likes(john,mary).
likes(paul,mary).
?- likes(john,X).
```

Part –B (5x3=15)

Write algorithms wherever required

6. Construct a decision tree for the given Table 6.1. (5)

Table 6.1 Data set

Weekend (Example)	Weather	Parents	Money	Decision (Category)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

7. Given the Table 6.1, Predict the class label for the given record X={Weather=sunny, Parents=No, Money=Rich} using Bayesian classification. (5)

8. Represent in First Order Logic (FOL) predicate clauses. (5)
   a. Two sets are equal if and only if each is a subset of the other.
   b. male and female are disjoint categories
   c. if two terminals are connected, then they have the same signal.
   d. Happy people are not sad.
   e. Caesar was a ruler
9. With a neat algorithm, explain Q-learning agent in detail. (5)

*** Do not wait to strike till the iron is hot; but make it hot by striking. - William Buell Sprague***

Department of Computer Technology, MIT Campus, Anna University

CP7253 - MACHINE LEARNING TECHNIQUES

2/4 MECSE – Assessment-II

Date: 12/04/17 Duration: 1.30 hrs Marks:50

Part-A (Answer any two) (2X10=20)

1. a. How nodes are classified based on best split(either it is pure or impure)
   b. Write an algorithm for k-nearest-neighbour classification given k and n, the number of attribute describing each tuples with example..
   c. Explain unidirected graphic model with example.

2. Write short note on;
   a) Compare Error estimation in bootstrap model with cross validation with example.
   b) Bagging estimation and boosting

3. Explain briefly about iterative optimization technique. Discuss in detail about an Expectation Maximization algorithm for Gaussian mixture model and maximum likelihood.

Part-B (Answer any two) (2X15=30)

4. Support vector machine work by finding a "maximum margin" separating hyperplane in a high-dimensional feature space.
   a) Why is the maximum margin hyperplane better then other separting hyperplanes?
   b) Support vector machines use a special trick for computing scalar product efficiently when fitting a hyperplane in a high-dimensional feature space. What is the trick?
   c) What is the relationship between the maximum margin hyperplane and a support vector?
   d) Consider a 2 class classification problem with a dataset of inputs $\{x_1 = \{-1, -1\}, x_2 = \{-1, +1\}, x_3=\{+1,-1\}, x_4 = \{+1, +1\}\}$ and a corresponding set of targets {t1, t2, t3, t4} where $t_i \in \{+1, -1\}$. Sing this feature space (no kernel trick), can we build a SVM to perfectly classify this dataset regardless of values of t_i's?

5. Load estimation

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	Young	False	False	Fair	No
2	Young	False	False	Good	No
3	Young	True	True	Good	Yes
4	Young	True	True	Fair	Yes
5	Young	False	False	Fair	No
6	Middle	False	False	Fair	No
7	Middle	False	False	Good	No
8	Middle	True	False	Good	Yes

Payouts and Probabilities: *[PO1]*

- Movie company Payouts
  - Small box office - $200,000
  - Medium box office - $1,000,000
  - Large box office - $3,000,000
- TV Network Payout
  - Flat rate - $900,000
- Probabilities
  - P(Small Box Office) = 0.3
  - P(Medium Box Office) = 0.6
  - P(Large Box Office) = 0.1

Part-B (Answer any three only) **(3 X 9= 27)**

6. i. How to design a learning system? *[PO3]*
   ii. Illustrate the designed learning system for handwriting recognition system. *[PO3]*

7. Discuss the least square approach for linear model for regression and its relation to maximum likelihood for minimizing the noise ratio relate to their target data set in detail. *[PO2]*

Car table database:

Color	Type	Doors	Tires	Class
Red	SUV	2	Whitewall	+
Blue	Minivan	4	Whitewall	-
Green	Car	4	Whitewall	-
Red	Minivan	4	Blackwall	-
Green	Car	2	Blackwall	+
Green	SUV	4	Blackwall	-
Blue	SUV	2	Blackwall	-
Blue	Car	2	Whitewall	+
Red	SUV	2	Blackwall	-
Blue	Car	4	Blackwall	-
Green	SUV	4	Whitewall	+
Red	Car	2	Blackwall	+
Green	SUV	2	Blackwall	-
Green	Minivan	4	Whitewall	-

8. Class label training tuples form all the given car table database *[PO2]*
   i) Compute the Gain ratio for the attribute "color" and "Type"
   ii) Compute the Gini index for the attribute "Type"

Customer Database Table:

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	Young	False	False	Fair	No
2	Young	False	False	Good	No
3	Young	True	True	Good	Yes
4	Young	True	True	Fair	Yes
5	Young	False	False	Fair	No
6	Middle	False	False	Fair	No
7	Middle	False	False	Good	No
8	Middle	True	False	Good	Yes
9	Young	True	True	Good	No
10	Middle	False	True	Excellent	Yes
11	Middle	False	True	Excellent	Yes
12	Old	False	True	Excellent	Yes
13	Old	False	True	Good	Yes
14	Old	True	False	Good	Yes
15	Middle	False	False	Good	Yes
16	Old	True	False	Excellent	Yes

9. i. What are Bayesian classifiers? Discuss the simple Bayesian classifier work flow.
   ii. Predicting a class label of a tuple using naive Bayesian classification from customer database table. The training data were shown in the table. The data table is described by the attribute Age, Has_Job, Own_Housing, and Credit_Rating. The class label attribute, Buy_House, has two distinct values, namely {c1=Yes, c2= No}. The tuple we wish to classify is Age=Middle, Has_Job=False, Own_House = True, Credit_Rate = Good. *[PO2]*